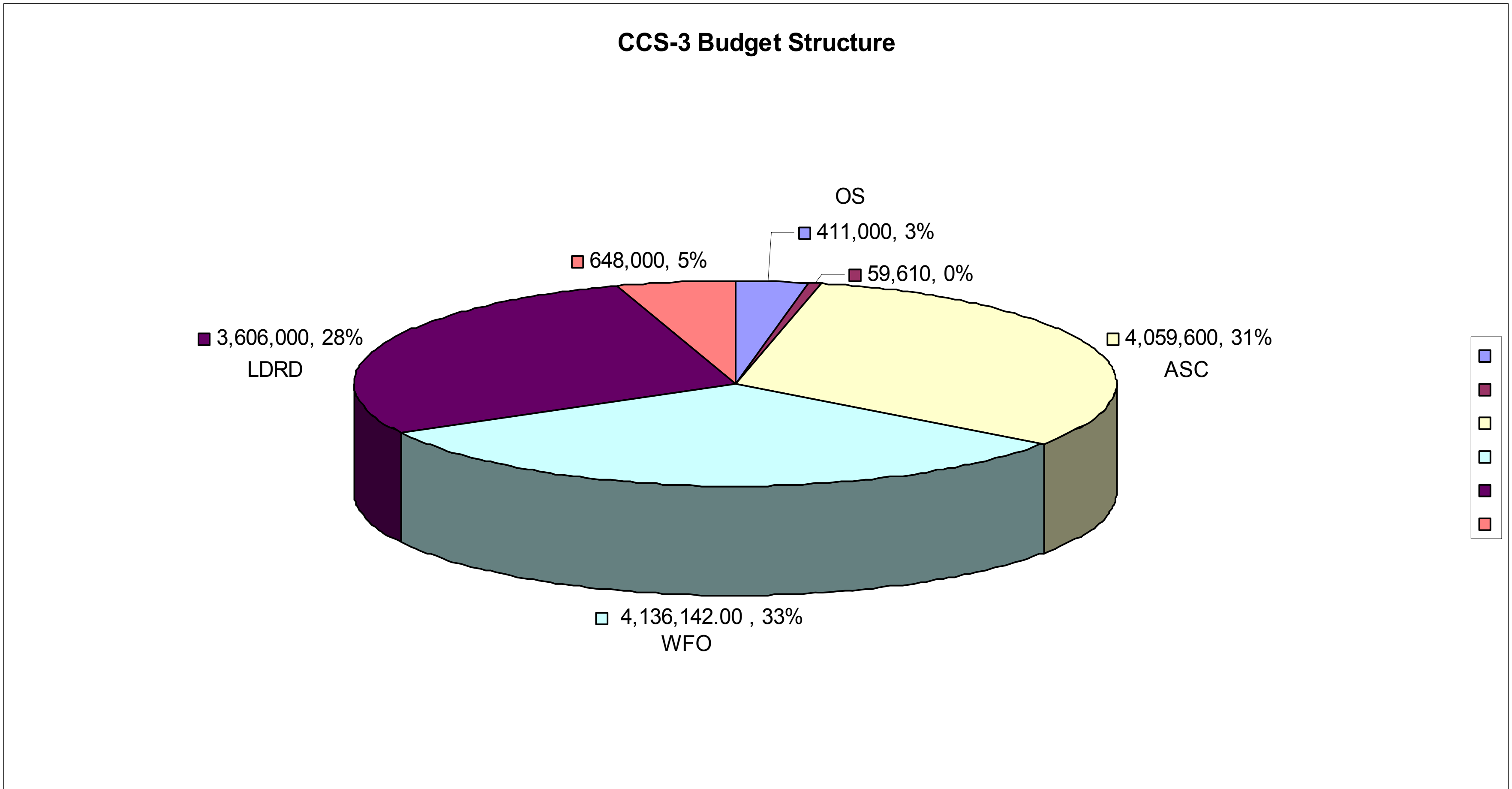
CCS-3 Modeling, Algorithms, and Informatics

- Computer Science and Mathematics Research
- Theoretical and Applied
  - High Performance Computing, Performance, Architecture
  - Machine Learning
  - Knowledge Systems and Computational Linguistics
  - Computational Biology and Bioinformatics
  - Multilevel Solvers
  - Quantum Information Processing

Tidbits

- 33 staff members, O(10) students, 4 postdocs
- Programmatic work in NW and TR
- Emphasis on scholarly work: in the year between DRC 2005 and DRC 2006, published 34 peer-reviewed journal papers, 20 peer reviewed conference proceedings papers, 3 edited books.
- Emphasis on software development in many areas
- Connected with various relevant scientific communities directly and through a myriad of internal and external collaborations
- O(50) charge codes…

CCS-3 Budget


CCS-3 Budget Structure
OS
411,000, 3%
59,610, 0%
648,000, 5%
3,606,000, 28%
LDRD
4,059,600, 31%
ASC
4,136,142.00 , 33%
WFO

Modeling, Algorithms, & Informatics (CCS-3)

Adolfy Hoisie, Group Leader
Francis J. Alexander, Deputy Group Leader
Erika Maestas, Staff Assistant

Margaret Tyler, Administrative Specialist
Justin Chavez, UGS Administrator
Hugh Greenberg, System Administrator

Performance & Architecture Lab (PAL)

Adolfy Hoisie, TL

Kevin Barker
Eitan Frachtenberg
Song Jiang
Greg Johnson
Darren Kerbyson
Mike Lang
Scott Pakin
Jose Sancho

Machine Learning & Pattern Recognition

Patrick M. Kelly, TL

Pat Fasel
Rich Fortson
John Hogden
James Howse
Don Hush
Olaf Lubeck
Reid Rivenburgh
Clint Scovel
Ingo Steinwart

Quantum & Classical Information Sciences

Marion Kei Davis, TL

Howard Barnum
Chris Brislwan
Leonid Gurvits
Paul Pedersen

Solvers

Michael Pernice, TL

Bobby Philip

Knowledge & Information Systems Science

Cliff A. Joslyn, TL

Karin Verspoor
Sue Mniszewski
Shou-De Lin
Ari Rosenberg

Information Physics & Modeling

Francis J. Alexander, TL

Marian Anghel
John Middleditch
Allon Percus

Computational Biology & Bioinformatics

Michael Wall, TL

Dengming Ming
Ilya Nemenman
Nikolai Sinitsyn

Performance and Architecture Lab (PAL)

- Research Areas:
  - High performance computing
  - Performance Analysis
  - Performance Modeling
  - System architecture
  - Interconnect networks
  - Scheduling
  - System software

- Recent Projects:
  - Performance of leading supercomputers, advanced architectures, IB, optical networks, system software performance, tool development for performance, OS scheduling for multi-cores, quantifying and exploiting communication-computation overlap, microprocessor architecture
  - ASC, DARPA, Office of Science

Performance and Architecture Lab: Recent Accomplishments

- Systems analyzed (measured and modeled):
  - Blue Gene/L (LLNL), RedStorm (SNL), Purple (LLNL)
  - ASC L2 Milestone on Blue Gene/L and RedStorm successfully completed
  - Multi-core analysis of AMD and Intel
- Workload modeled increased
- Novel Networks explored:
  - Optical Circuit Switched network (OCS) with IBM, SC05
- Novel systems analyzed:
  - IBM PERCS, HPCS, multitude of configurations explored, used in HPCS review and Phase 3 proposal
  - Accelerators – case study of S_N transport on Clear-speed CSX600 completed
- World-class publication record in 2005/6:
  - 14 Journal Papers, 20 Conference Papers, and 2 edited books,
  - includes 3 in IEEE Transactions (Computers and Parallel & Distributed Processing)
- Installation of 1024 core AMD system with Infiniband
  - For performance analysis, development, and software stack analysis

Quantum and Classical Information Sciences

Research Areas:

- Quantum Information Processing
- Theory of Multirate Digital Filter Banks and Wavelet Transforms

Example Applications:

- Quantum and Classical Cryptography
- Quantum and Classical Complexity
- Quantum Computing
- Quantum Physics
- Digital Image Source Coding
- RF Remote Sensing

Recent Contributions

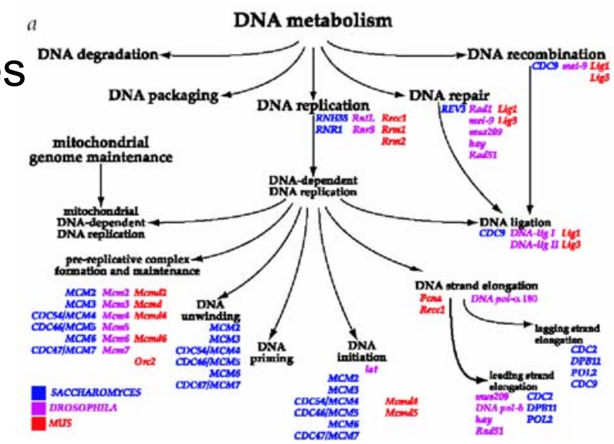
- Settled long-standing conjecture on NP-hardness of checking whether a given bipartite mixed quantum state is entangled
  - Implications for QIP, signal processing, and computer vision
- Quantification of robustness of quantum entanglement
  - Implications for both theory, and practical implementation of quantum cryptography and computing
- Found deep connections between fundamental classical complexity problems and quantum entanglement.
  - May lead to tractable, deterministic classical algorithms for a large class of problems solvable by quantum computing
- Proposal for experimental continuous-variables quantum key distribution
  - May greatly increase secret key rate by using telecom-industry technology rather than single-photon states and detectors

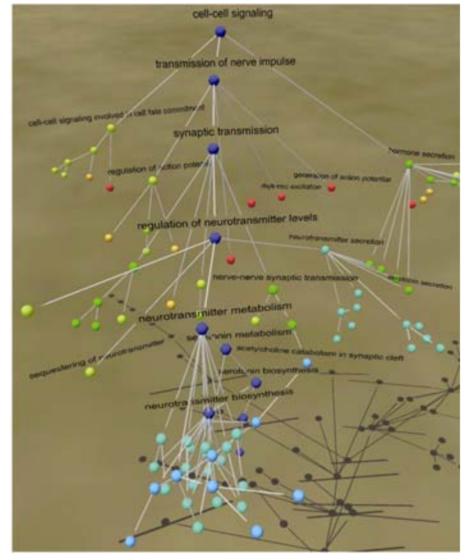
Knowledge and Information Systems Science

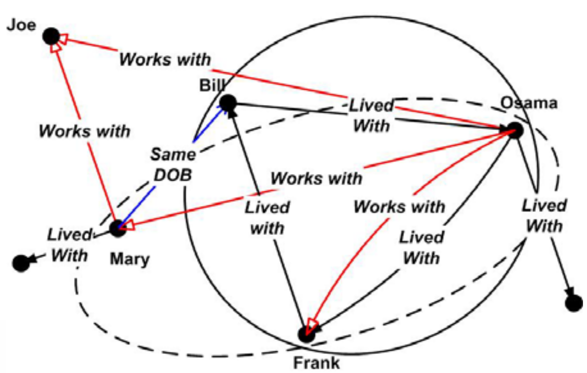
- **Knowledge discovery and extraction:**
  - Emphasis on *semantic* information
  - *Hybrid approach* combining statistical, numerical, and quantitative with symbolic, logical, and qualitative techniques
- **Ontologies and Conceptual Semantic Systems:** Discrete mathematical, combinatorial, and order theoretical approaches
- **Computational Linguistics:** Text extraction, knowledge discovery, search and retrieval, corpus management
- **Database Analysis:** User-guided knowledge discovery in complex, multi-dimensional data spaces
- **Software Architectures:** Parallel and high performance algorithms for knowledge systems, simulations

Ontologies and Knowledge Representation

- **Bio-Ontologies**
  - Discrete math approaches to large semantic hierarchies
  - Applied lattice theory
  - Information visualization (D-4 collaboration)
  - For protein function inference







- **Semantic Network Databases**
  - Statistical techniques for typed path analysis
  - Abnormal node and connection discovery
  - Intelligence analysis and homeland security

Computational Linguistics and Text Extraction

- **Los Alamos Semantic Event Recognizer (LASER)**
  - Uses templates for relation extraction from text
  - Unsupervised and weakly supervised machine learning techniques for template creation

In 3T3-L1 adipocytes, a low level of serine phosphorylation of stat3 on residue 727 was observed and was markedly enhanced by ins or os.

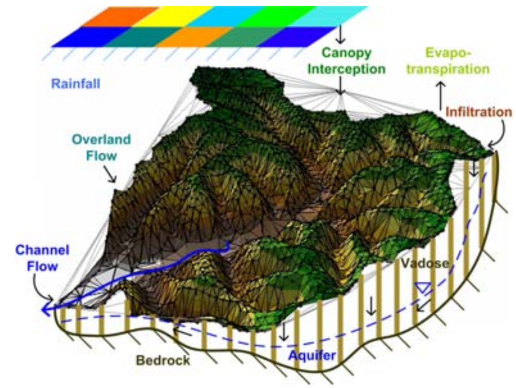
ins	**phosphorylate**	**stat3**
os	**phosphorylate**	**stat3**

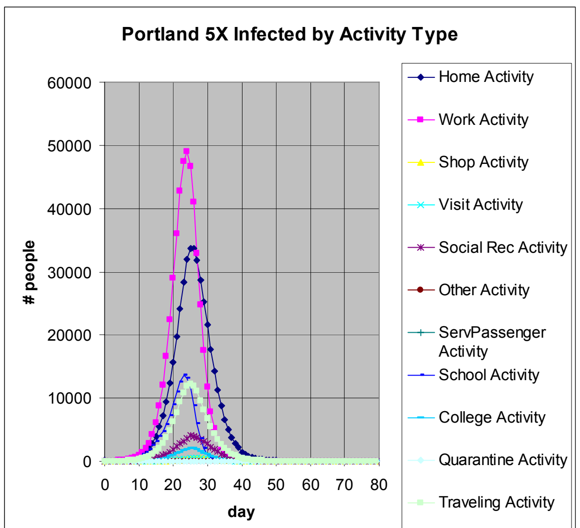
- **Extraction of Motivation and Intent from Text**
  - Statistical content analysis of phrases annotated by affect, mood, or emotion
  - In interaction with syntactic/phrasal structure

[The Party]$_{subject}$ [denounces]$_{verb}$ [[the terrorizing of innocent people] and [the use of emergency law [against honest citizens]]]$_{object}$.

Parallel and High Performance Scientific Algorithms

- **High Resolution Physically-Based Model of Semi-Arid River Basin Hydrology**
  - Parallelization of watershed models for used for validation and prediction
  - Institutional Computing resources, TLC, Lambda, and Mauve

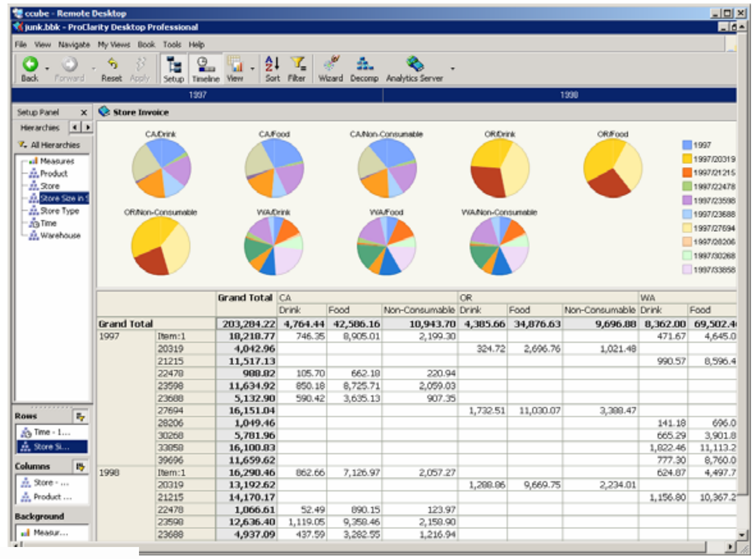




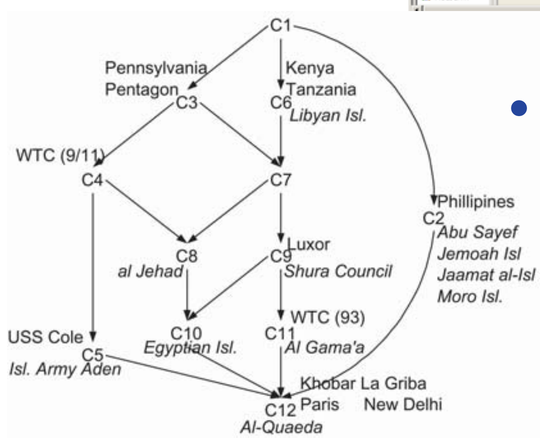
- **Epidemic Simulation System (EpiSimS)**
  - Public Health Module of NISAC Urban Infrastructure Suite (UIS)
  - Simulation of metropolitan disease spread reflecting social contacts and disease transmission

Cubic Lattice Data Analysis

- **Database hypercubes**
  - User-guided knowledge discovery in high-dimensional databases
  - OnLine Analytical Processing (OLAP)
  - Cubic lattice formulation



	Kenya	Phillipines	WTC 93	La Griba	Pennsylvania	Cole	Khobar	Luxor	WTC 9/11	Tanzania	Paris	New Dheli	Pentagon
Al Qaeda	X	X	X	X	X	X	X	X	X	X	X	X	X
Eg. Islammic J	X				X			X	X	X			X
Abu Sayyaf		X											
Jemmah Islamiah		X											
Shurah Council	X				X			X		X			X
Isl. Army of Aden					X	X			X				X
al Jehad	X				X				X	X			X
Al-Gama'a al Isl.	X		X		X			X		X			X
Libyan Isl. Group	X									X			
Jaamat al-Isl.		X											
Moro Isl. Lib Front		X											



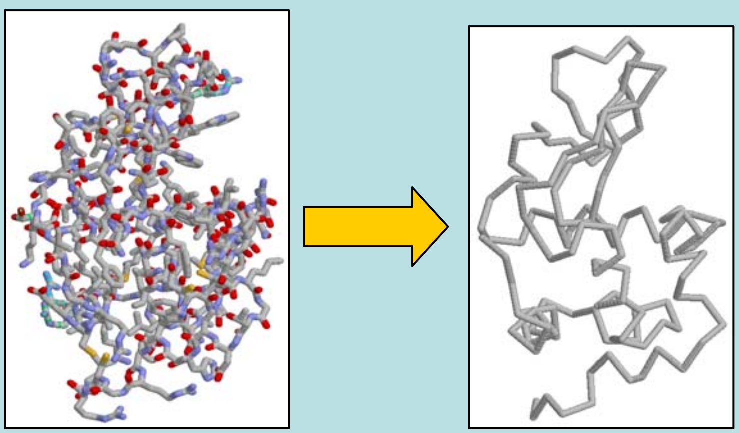
- **Formal Concept Analysis**
  - Extracting hierarchical structure of relational data
  - Ontology induction

Computational Biology and Bioinformatics

- Capabilities
  - Mathematical Analysis and Modeling
  - Statistical Physics
  - Complex Systems and Nonlinear Dynamics
  - Information and Learning Theory
- Applications
  - Genetic Regulatory Networks
  - Biomolecular Interactions
  - Protein Structure and Function
  - Neuroscience
- Projects
  - LDRD-DR (Wall, PI) “Computational Methods for Protein Function Inference”
  - LDRD-ER (Wall, PI) “Design Principles of Genetic Regulatory Networks”

Coarse-Grained Modeling of Proteins

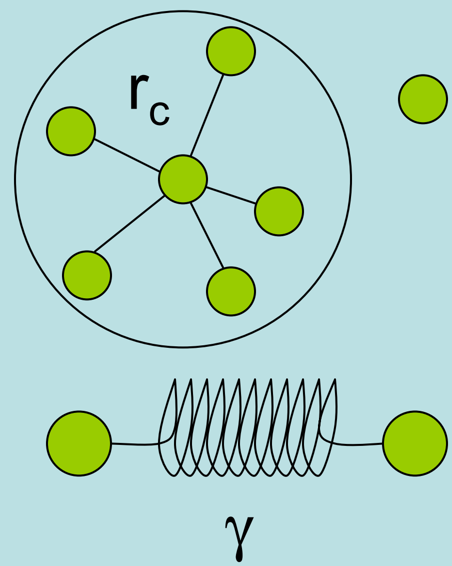
- Extract key properties of detailed models for use in large-scale simulations
  - Molecular signaling mechanisms
  - Free-energy estimation
- Make simulation of large macromolecular systems feasible
- Example: Elastic Network Model



All-atom model

~10x-reduced model
(select alpha-carbons)

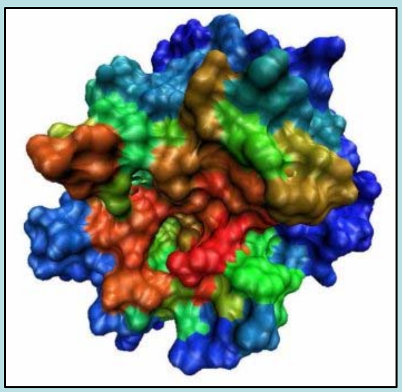
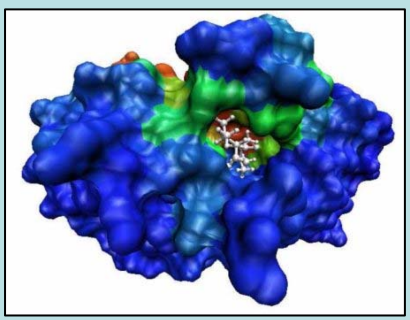
Tirion. PRL 1996
Atilgan et al. Biophys J 2001



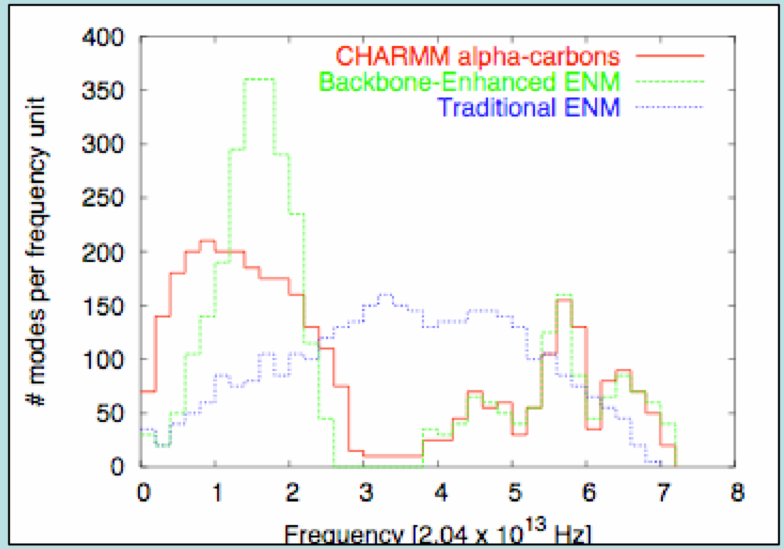
Network of interacting atoms

Improved Coarse-Grained Model

- Discovered a problem with the elastic network model
  - Incorrect density-of-states
- Fixed the problem
  - Strengthen backbone interactions
- Applied the new model to analyze communication between binding sites in proteins

Binding at one site (left) causes a large change in the second site (right)



Unimodal density of states (blue) disagrees with all-atom model (red)

Bimodal distribution is recovered by strengthening interactions between backbone neighbors (green)

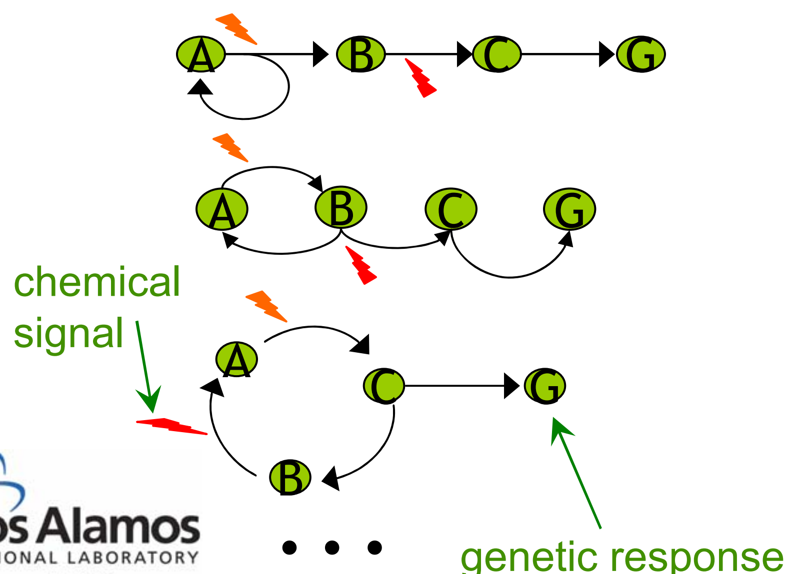
D. Ming & M.E. Wall. *Phys Rev Lett* 95 (2005) 198301

Information processing in biochemical circuits

- How many bits can be transmitted by a realistic stochastic biochemical network? How many different cell fates can be encoded?
  - With constrained molecular copy number?
  - With constrained time delays?
- How does this number depend on the network topology? (Are there special topologies?)
- Is this number robust to parameter fluctuations?

Analyze all topologies

exactly one promoter per gene,
each TF binds to one promoter type



Analyze steady states *g=g(c)* for all signals

$$\frac{dg}{dt} = -R_g g + a_0 + \alpha\left(\{g, c\}\right) = 0$$

…

$$\theta$$

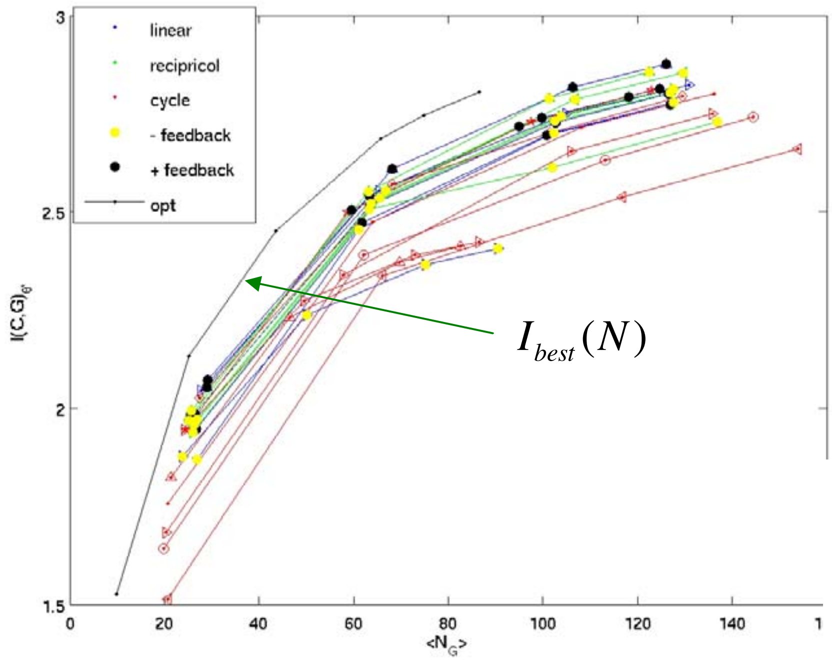
Add Langevin noise and evaluate

$$\hat{\theta} = \arg\max_\theta I(C;G\,|\,\theta) - \lambda_1 N - \lambda_2 T$$

$$I_{\max} = I(C;G\,|\,\hat{\theta})$$

COMPUTER & COMPUTATIONAL SCIENCES

All networks achieve (almost) maximum MI

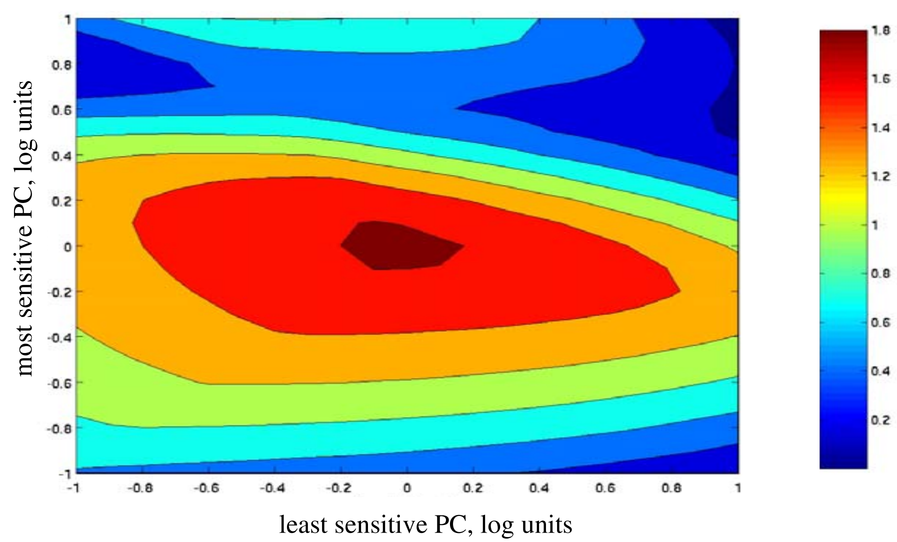


$$T \leq \text{hours}$$

All topologies achieve mutual info close to the maximum possible (e.g., at *N*=80 the max is 2.7 bits, while *the worst* topology allows about 2.4 bits.

Is MI sensitive to parameters?

- Almost 10-fold parameter changes may still lead to
  $$I \geq 0.8 I_{best}$$
- High *I* is generic! No fine-tuning.



Important Research Directions

- Computational methods for protein function inference
- Reverse engineering of biological networks
- Information processing in biological networks
- Mathematical modeling of gene regulation
- Coarse-grained modeling of biological systems
- Free-energy estimation for protein interactions
- Neural coding and animal learning
- Intelligent computing

Michael Wall / mewall@lanl.gov

CCS-3 Solvers Team

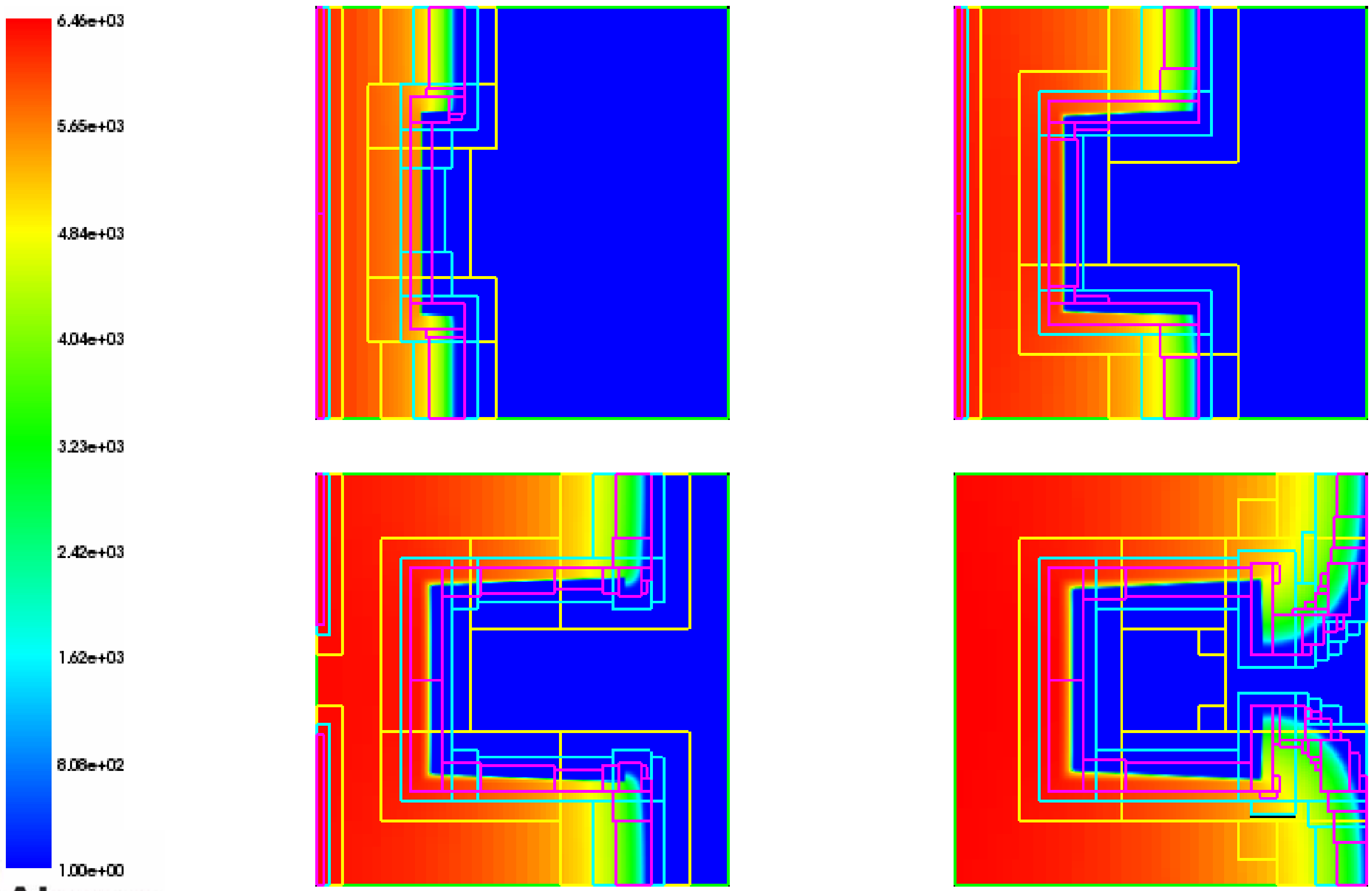
Capabilities:

- Full spectrum of linear/nonlinear solver methods
  - Multigrid, Krylov subspace methods, domain decomposition, Newton-Krylov methods
  - Special focus on optimal multilevel solvers for AMR grids
    - FAC, AFAC, AFACx methods as preconditioners
- Use with implicit time integration
- Adaptive mesh refinement
  - Grid management frameworks (*SAMRAI*)
  - Feature-based refinement criteria
  - *Verified* second-order spatial discretization schemes
- Ancillary software products
  - *SAMRSolvers*, *SAMRUtilities*
  - Solver software interoperability (*PETSc*, *KINSOL)*

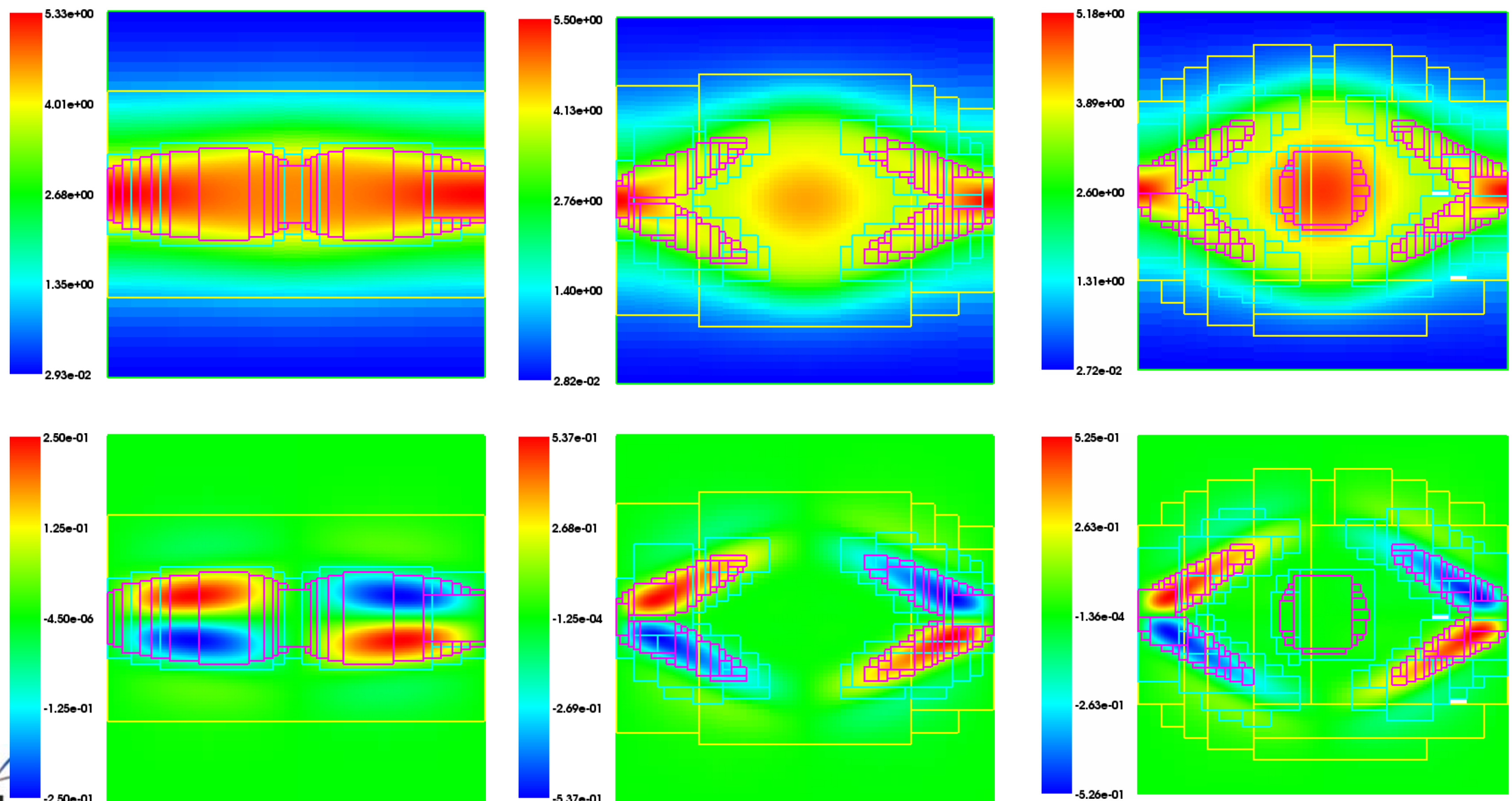
CCS-3 Solvers Team Projects

- Application of implicit AMR to Marshak wave problem
  - Produces same solution with 80-90% *fewer* mesh points
  - AMR up to *10x faster* than equivalent fine grid calculation
- Application of implicit AMR to reduced resistive MHD
  - Goal: enable study of magnetic reconnection for low resistivities at unprecedented spatial resolution
  - First-of-a-kind simulation of tearing mode case
    - Showed weak dependence of iteration counts on mesh configuration
  - Supported by LDRD
- Comparison of FAC, AFAC with AMG on AMR grids
  - Use Marshak wave example
  - Try to quantify potential impact of future programmatic investment in geometric methods
  - Supported by WSR

Marshak Wave Sample Calculation



Tearing Mode Sample Calculation



Machine Learning Team

Research Areas:

- Computational Learning Theory
- Data Driven Modeling
- Statistical Inference
- Pattern Recognition
- Anomaly Detection
- Neural Networks
- Digital Image/Signal Processing
- Image Understanding
- Clustering Algorithms
- Data Analysis

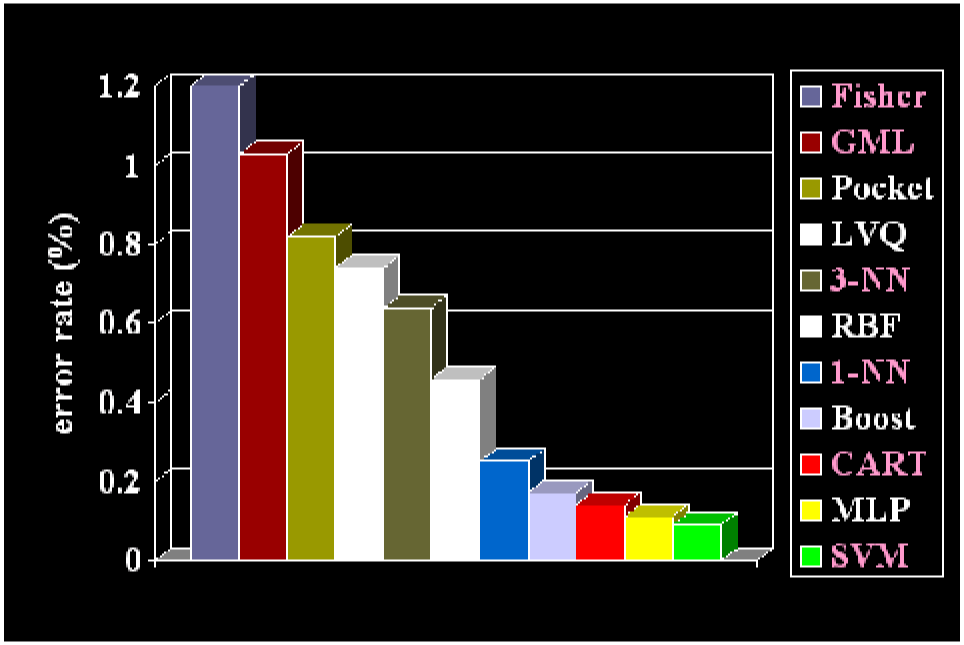
Example Applications:

- Satellite Image Analysis / Automated Tracking of Urban Growth Trends
- Fraud Detection (Tax Returns, Medicare Submissions)
- OCR Improvement for Degraded Document Images
- Automated Language Identification from Document Images
- Speech Recognition

Basic Research in ML (Recent Contributions)

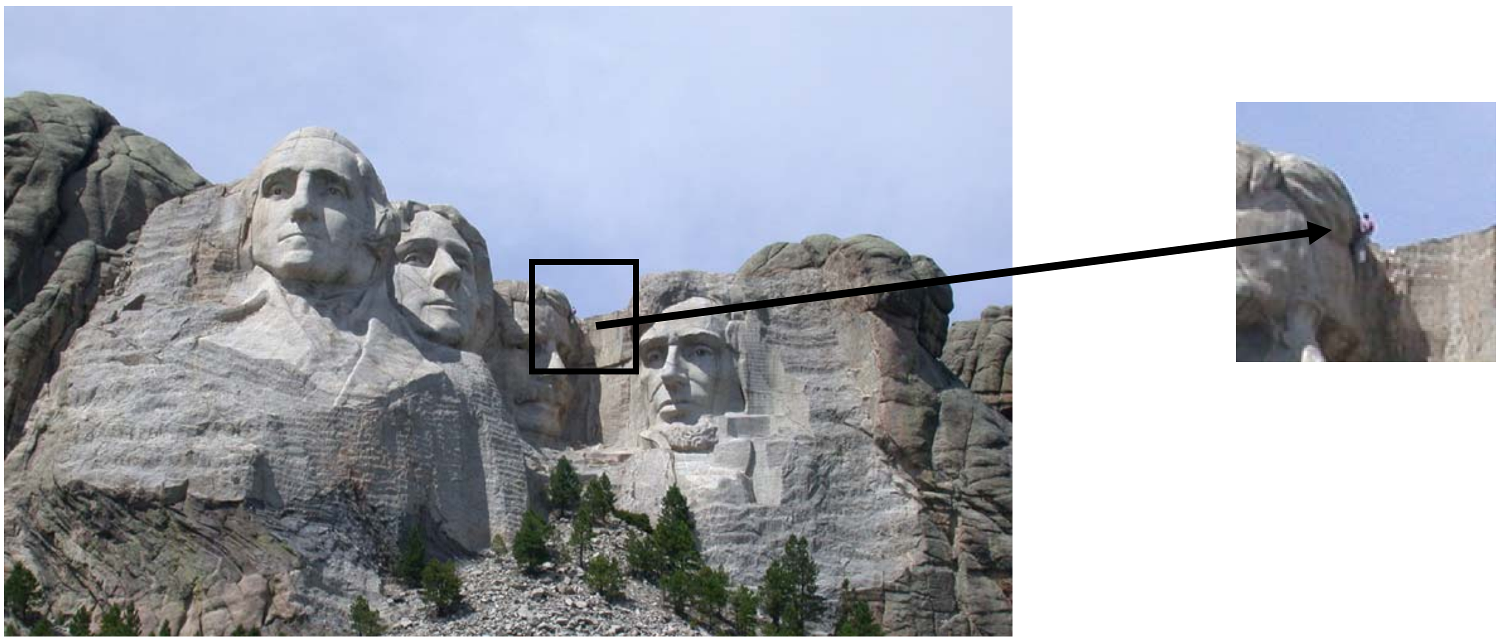
- **First Complete Solution Method** for the generalized version of the supervised classification problem
  - Near-Optimal Performance is Guaranteed
  - Run-Time Guarantees that are both Feasible & Efficient

DARPA Intrusion Detection Data Set



Basic Research in ML (Recent Contributions)

- **First Practical Anomaly Detection Framework** that provides a method for validating anomaly detector performance



Current Image Analysis Efforts

- Image Comparison Methodologies (FEEMADS)
  - Simulated Data vs. Experimental Results
  - Contributed validation techniques for three separate Level 2 ASC milestones
- Image Analysis / Pattern Recognition
  - "Find me all of the other things in the image that look like this…"

Information Physics and Modeling Team

Research Areas:

- Variational Modeling of Large-Scale Nonlinear Systems
- Multiscale / Hybrid Numerical Methods
- Statistical Physics of Algorithms
- Statistical Physics (Foundations)
- Nonlinear Prediction
- Computational Kinetic Theory/Fluid Dynamics
- Accelerated Monte Carlo Methods
- Optimal Estimation / Inverse problems

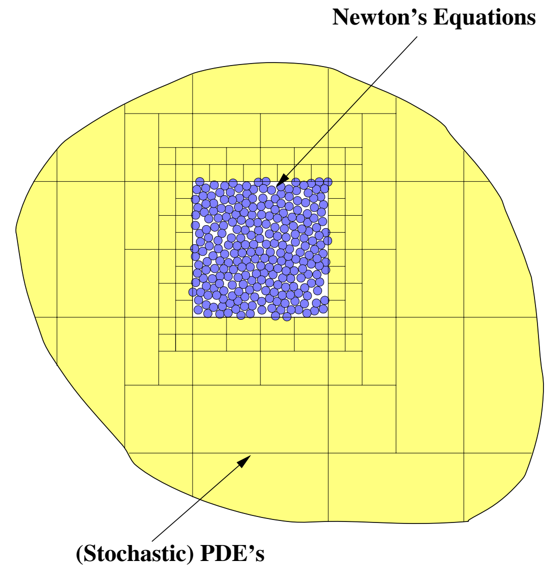
Applications:

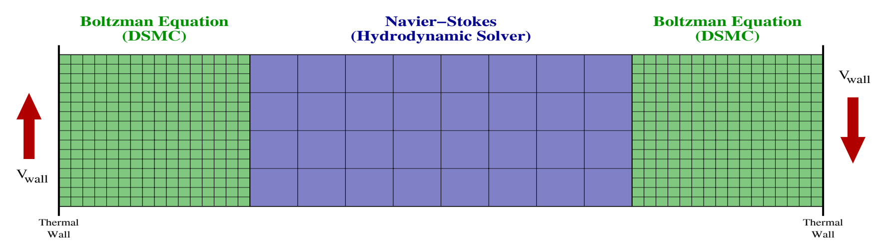
- Non-convex optimization
- Phase Transition Kinetics and Dynamics
- Earthquake / Fault Modeling
- Powergrid Modeling and Analysis
- Source Identification /History Reconstruction
- Fast Langevin Molecular Dynamics
- Pulsar Astrophysics “Starquakes”

Hybrid Methods

*The accurate and efficient simulation of multiscale phenomena**

- Different spatial regions require different physics/resolutions
- Use computationally cheapest, ***valid***, method in each region
- Dynamically couple methods
- Computational Gain $=V_{macro}/V_{micro}$

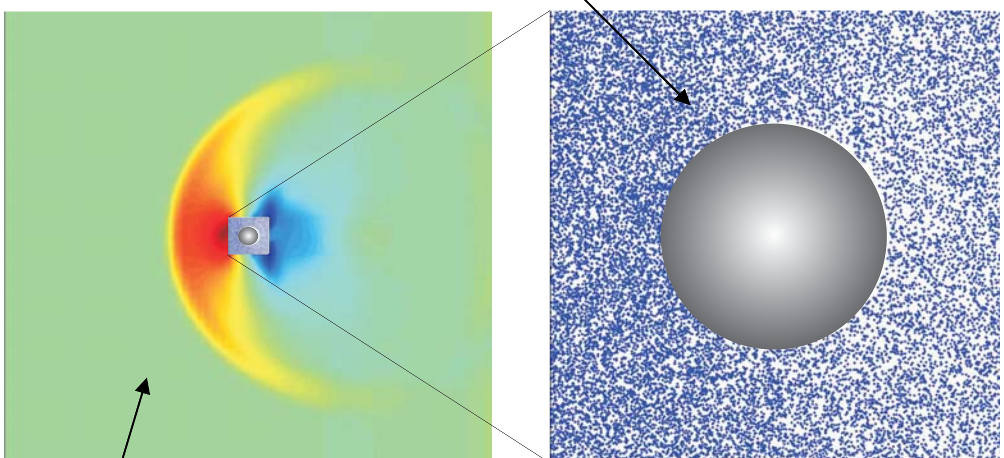




Hybrid Methods

Boltzmann Equation

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_r f + \frac{\mathbf{F}}{m} \nabla_v f = \Omega_{coll}$$



Euler / Navier-Stokes

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$$

$$\frac{\partial \rho \mathbf{u}}{\partial t} + \nabla \cdot \mathbf{\Pi} = 0$$

$$\frac{\partial \rho e}{\partial t} + \nabla \cdot (\rho e \mathbf{u}) + \nabla \cdot \mathbf{q} + \mathbf{P} : \nabla \mathbf{u} = 0$$